修 士 学 位 論 文

Master's Thesis

Distinguishing conspecific bats by their echolocation calls using a convolutional neural network

同志社大学大学院　生命医科学研究科

医工学・医情報学専攻 医情報学コース

博士課程（前期課程）2021 年度 5503 番

Doshisha University

Graduate School of Life and Medical Sciences

Technische Universität Wien

E363 Institute of Biomedical Electronics

UE 066 453 – Master's Programme Biomedical Engineering

Matrikelnummer: 01525261

Academic advisor: Univ.Prof. Dipl.-Ing. Dr.techn. Eugenijus Kaniusas

氏 名 / Name: Andreas Mayer

指導教員 飛龍 志津子 教授

Under the supervision of

Prof. Shizuko Hiryu

June, 2023

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Kyōto, Juni 2023

Andreas Mayer

Statutory Declaration

I declare that I have authored this thesis independently and performed the associated research myself, using only literature cited in this volume. If text passages from sources are used literally, they are marked as such. I confirm that this work is original and has not been submitted elsewhere for examination, nor is it currently under consideration for a thesis elsewhere.

Kyōto, June 2023

Andreas Mayer

Kurzfassung

Unterscheidung konspezifischer Fledermäuse durch ihre Echolokationsrufe mittels eines Convolutional Neural Networks

von Andreas Mayer

Fledermäuse verwenden Echolokationsrufe um ihre Umgebung wahrzunehmen. Folgendermaßen wird in diesbezüglichen Forschungsarbeiten in der Natur oft auf Audioaufnahmen zurückgegriffen. In vergangenen Studien wurde gezeigt, dass die Ortungsrufe zur Diskriminierung zwischen unterschiedlichen Spezien genügen, jedoch gab es bisher keine Versuche mittels Echolokationsrufen zwischen Individuen der selben Spezies zu unterscheiden. In dieser Arbeit zeige ich, dass es bei 2 Spezien (*Miniopterus fuliginosus*, *Pipistrellus abramus*), durch Anwendung eines 4-lagigen Convolutional Neural Networks (CNN), welches mit Spektrogrammen von Fledermausrufen trainiert wurde, möglich ist, konspezifische Individuen mit einer hohen Erfolgsrate zu unterscheiden. Die Fledermäuse wurden im Einzel- und Paarflug aufgenommen, ihre Ortungslaute gesammelt und als Datenbank genutzt, um verschiedene Modelle auf Basis eines Neuronalen Netzwerks (NN) zu trainieren und evaluieren. Die Modelle selbst wurden mit Local Interpretable Model-Agnostic Explanations (LIME) analysiert, einer Methode die erklärt, aus welchem Grund das NN bei seiner Entscheidung ankommt. Die F1 Werte der Klassifikationen über alle Experimente reichen von 0,710 bis 0,983. Diese Ergebnisse weisen auf eine individuell spezifische Signatur in den Echolokationsrufen hin. Die Analyse mittels LIME legt nahe, dass diese die Endfrequenz der absteigend frequenzmodulierten Rufe ist. Änderungen der Klassifikationsperformance des NN deuten auf ein Jamming-avoidance Verhalten der Fledermäuse hin.

Abstract

Distinguishing conspecific bats by their echolocation calls using a
convolutional neural network

by Andreas Mayer

Bats use echolocation calls to perceive their surroundings. Consequently, sound measurements
are often used to study them in the field. It has been shown that their calls are enough to identify
the species of bat reliably, but there have been no attempts to use sounds to discriminate between
individuals within the same species. In my thesis, I show that with 2 separate species
(*Miniopterus fuliginosus*, *Pipistrellus abramus*) it is possible to identify conspecific individuals
with a high success rate, when using a 4 layer convolutional neural network (CNN), that is trained
on spectrograms of bat echolocation calls. The bats were recorded while flying alone and in pairs,
with the calls collected being used as a database to train and evaluate NN models. The models
themselves were analyzed using Local Interpretable Model-Agnostic Explanations (LIME), a
method showing how the NN arrives at its prediction. F1 scores for classification across all
experiments ranged from 0.710 to 0.983. These results indicate an individual specific cue within
the echolocation calls, which with LIME was found to likely be the terminal frequency of the
downward frequency-modulated calls. Change in classification performance of the NN in pair
flight indicates jamming avoidance behaviour is employed by the bats.

Contents

1 Introduction

1.1 Bat Echolocation

While most mammals use vision to navigate their environment, bats use echolocation to perceive their surroundings. With their acoustic imaging system they can not only determine the location of an echo source, but also perceive its size, form, and surface texture. Since the bats produce the carrier signal, an, in general, ultrasonic pulse, themselves with their mouth or nose, their navigational system can be described as an active orientation system. [1]

The advantage of being independent from sunlight is offset however, by a limited duty cycle, sound field, range and resolution, as well as relatively high energy cost per pulse emitted. Evolutionary it was therefore important for each of the over 1100 bat species [2], most of which utilize echolocation, to be as efficient as possible in their respective environment and adapt how they produce ultrasonic pulses and listen for the corresponding echoes.

Typically, bats use ultrasonic, downward frequency-modulated (FM) pulses as their echolocation calls (fig. 1, left). However, there are many varieties of call structures. Some species, for example those of the family *Rhinolophidae*, have adapted their hearing to a very narrow frequency band and consequently produce calls with a constant frequency (CF), which additionally may have FM parts before and/or after the CF one (fig. 1, center). Furthermore, other species may first use steep downward frequency sweep which is then followed by a quasi-constant frequency part (FM-QCF, fig. 1, right) where the frequency gradient approaches zero. The final, and also lowest, frequency reached in these calls is called the terminal frequency (TF) of the call.

Due to the vocal anatomy needed to emit these pulses, in CF bats the largest amount of energy is usually present in the second harmonic of the call, while the first harmonic is attenuated by the nasolaryngeal tract. In FM emitting bats, the fundamental frequency is the most strongly emitted mode. [1]
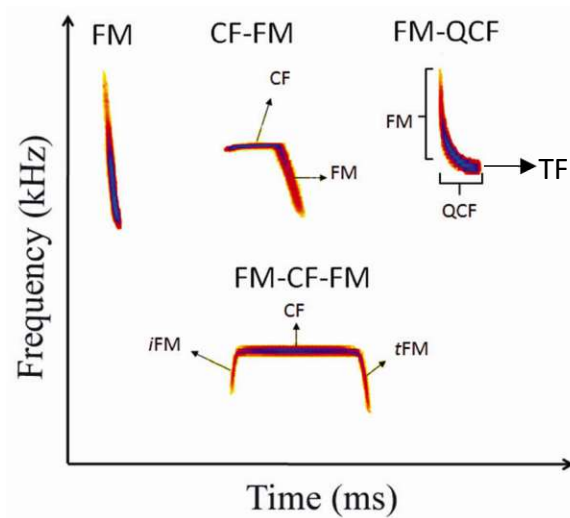
Figure 1: Types of echolocation calls in bats. Abbreviations: FM = frequency modulated, CF = constant frequency, QCF = quasi constant frequency, TF = terminal frequency. Figure modified from [3]

1.2 Echolocation call categorization

Several studies have been published that automatically categorize bat calls recorded from different species. Through machine-learning algorithms theses calls could be attributed with high accuracy to a certain bat species of the respective database that was used to train the used model. [4]–[8] One of the motivations for this approach is to be able to monitor the activity and quantity of bat populations in the wild, as well as which species are present in a certain location. [9] This could be useful not only for understanding the animals better but also to protect them through conducting environmental impact assessments employing this method.

So far, however, these systems are missing an important factor when evaluating the presence of bats: the number of individuals that make up the population. An automated system that counts individuals in the wild would need to differentiate between calls of not only different species, but bats of the same species at the same time. Further applications of distinguishing conspecifics are monitoring feeding behaviour and to understand the communicative potential of the calls better. Moreover, a system that is able to distinguish between conspecifics that fly simultaneously, might help in research regarding the jamming avoidance behaviour of bats. This phenomenon has been shown in bats flying while surrounded by a complex auditory scene, for example other bats that also emit echolocation calls. It has been discussed, that in this situation bats will change parts of their calls, presumably the TF, so they can distinguish their own echoes from those of other individuals more easily. [10]

This thesis describes the development and results of a classifier based on a convolutional neural network (CNN), used to distinguish between spectrograms of echolocation calls of individual bats of the same species in two different species.

After individual distinction, LIME (Local Interpretable Model-Agnostic Explanations) [11], an explanation technique for the classifications made by neural networks, is used to get an insight into which parts of the bats' calls is important for categorizing them to an individual. In addition to helping future work on a system that counts individuals and elucidates jamming avoidance behaviour, these explanations may give insights into which call features are important for the bats themselves to distinguish between each other. [12]–[16]

2 Methods

2.1 Bat species

Two bat species, *Miniopterus fuliginosus* (Hodgson, 1835) and *Pipistrellus abramus* (Temminck, 1840), that use downward FM-type echolocation pulses with a quasi-constant terminal frequency (QC-TF) part, were investigated. Figure 2 shows an example of echolocation pulses of both species. During flight, bats will vary the frequency bandwidth and pulse duration in response to the surrounding conditions.
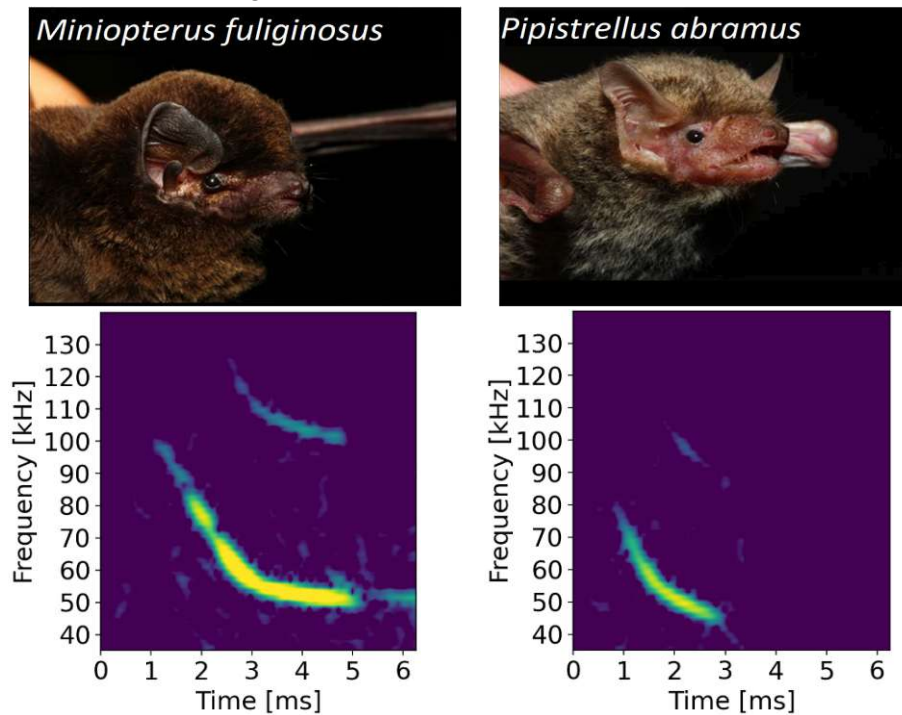
Figure 2: Representative spectrograms of individual echolocation calls of the bat species used.

The bats were housed in the laboratory of the Faculty of Life and Medical Sciences and were kept on an ad libitum nutrient-enriched diet based on live mealworms (larvae of *Tenebrio molitor*) and vitamin-enriched water. *P. abramus* individuals were kept in rearing cages [25 (L) × 17 (W) × 17 cm (H)] in groups of 3 to 5 individuals at $25 \pm 2°C$ and 55 ± 5 % relative humidity, and with the natural light-dark cycle of the photoperiod in Kyotanabe, Japan. *M. fuliginosus* individuals were kept together in a flight room [4 (L) × 3 (W) × 2 m (H)] that enabled them to fly regularly; at $20.5 \pm 1°C$ and 90 ± 5 % relative humidity. The day-night cycle of the room was set to 12 h:12 h dark:light.

M. fuliginosus individuals were captured in Fukui prefecture in a forest close to the Disaster Prevention Research Institute with permission by the Fukui Prefectural Government, Natural Environmental Division. They have been shown to change the TF part of their calls when exposed to FM jamming sounds mimicking other bats, shifting their own TF upwards by up to 2 kHz. [10]

This is called the jamming avoidance response and may help individuals flying in groups to listen to the echoes they have produced themselves and not by mistake to those from other individuals. Furthermore bats themselves may distinguish each other by listening specifically for the TF part of their conspecifics' echolocation calls. [10], [14], [17], [18] This species' echolocation calls sweep downward from about 100 kHz to about 50 kHz, where their hearing sensitivity peaks at frequencies between 44 and 56 kHz. [19]

P. abramus individuals were captured at a roost in Kyotanabe, Kyoto prefecture. They have been shown to increase their TF by even more than 2 kHz when exposed to jamming sounds with frequencies below that of their own pulses. [20] Furthermore they will decrease the TF when foraging, in a trade-off for a wider sonar beam-width [21], and adapt their call intensity to have a constant echo intensity [22] as well as reduce the duration of their calls when approaching prey. [23] This species' echolocation calls sweep downward from about 80 kHz to about 40 kHz, where their hearing sensitivity peaks at frequencies between 35 and 50 kHz. [24]

2.2 Audio recording

As a first step in training a neural network to classify calls, suitable training data had to be obtained. Different recording setups were applied. First, *M. fuliginosus* and *P. abramus* bats were recorded while flying alone in a flight chamber (fig. 3). In these 7 sessions a standing microphone (Avisoft CM16), sampling at 500 kHz, in conjunction with an analog amplifier (DT MA3) and analog bandpass filter (15 kHz – 150 kHz, NF 3625) were used.
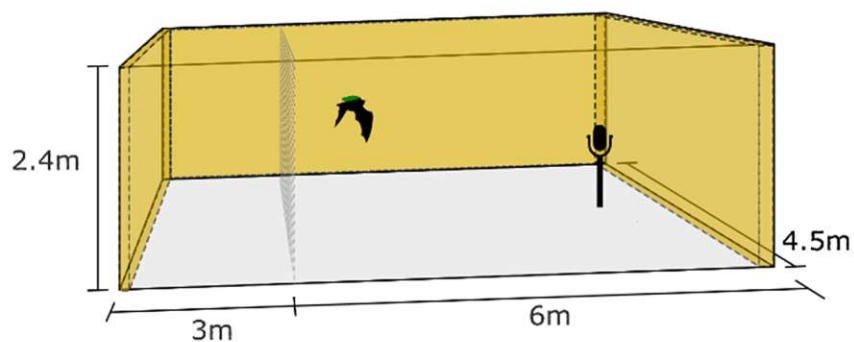

Figure 3: Flight chamber with standing microphone. Bats fly freely in a flight space separated by a net, sizing 4.5×6×2.4m.

Secondly, pairs of *M. fuliginosus* were flying together at the same time. For this setup it was not longer enough to just use a standing microphone, as the data would need to be correctly labelled (i.e. which call originated from which respective bat) and it would be impossible to tell apart the origin of each call later. Therefore, a custom-made, on-board recording device (data logger) was

developed in cooperation with Girlier company during the course of this research. It records with a 288 kHz sampling rate, at 16 bit (96 dB dynamic range) and uses an omnidirectional MEMS (microelectromechanical systems) microphone. With an 50 mAh (nominal value) battery the typical maximum recording length was 3-5 minutes.

Using this device is also the reason why the pair flight experiment could only be conducted with *M. fuliginosus*: even though the weight of the logger is light, it is still more than 40% of a *P. abramus* individual's bodyweight, prompting them not to fly when it was attached (table 1).

Table 1: Bat weight compared to data logger weight.

Bat species	typical weight [g]	Data logger weight (abs., including 50 mAh battery) [g]	Data logger weight (rel. body weight)
M. fuliginosus	13.0	2.9	22 %
P. abramus	6.5	2.9	44 %

The data logger's design was updated during the course of the study as the first version intermittently skipped recording for a few seconds, making synchronization impossible. Furthermore the weight was reduced in the second version. The logger was attached to the back of the bat by using a strip of double-sided tape (fig. 4).
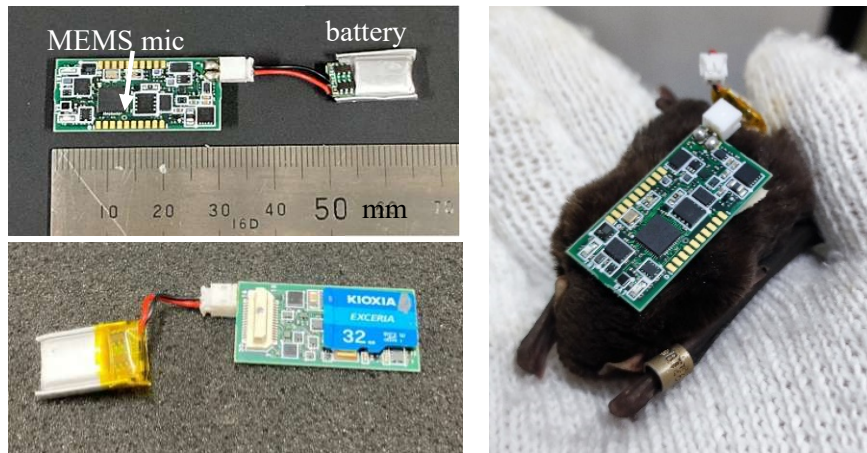

Figure 4: The first (left, top) and second (left, bottom) version of the on-board data logger. *M. fuliginosus* with the first version of the data logger attached on its back (right).

The MEMS microphone records with high sensitivity and low noise in the typical frequency ranges of both bat species used (fig. 5). Since it was attached to the bat's during flight, the recordings made this way were free from any Doppler shifts and sound attenuation, as when recording with a standing microphone, and true to what the bats could hear themselves.

To have an accurate comparison of calls emitted during solo and pair flights, all bats that flew in pairs were also recorded with the logger while flying alone to remove the microphone used as a variable.
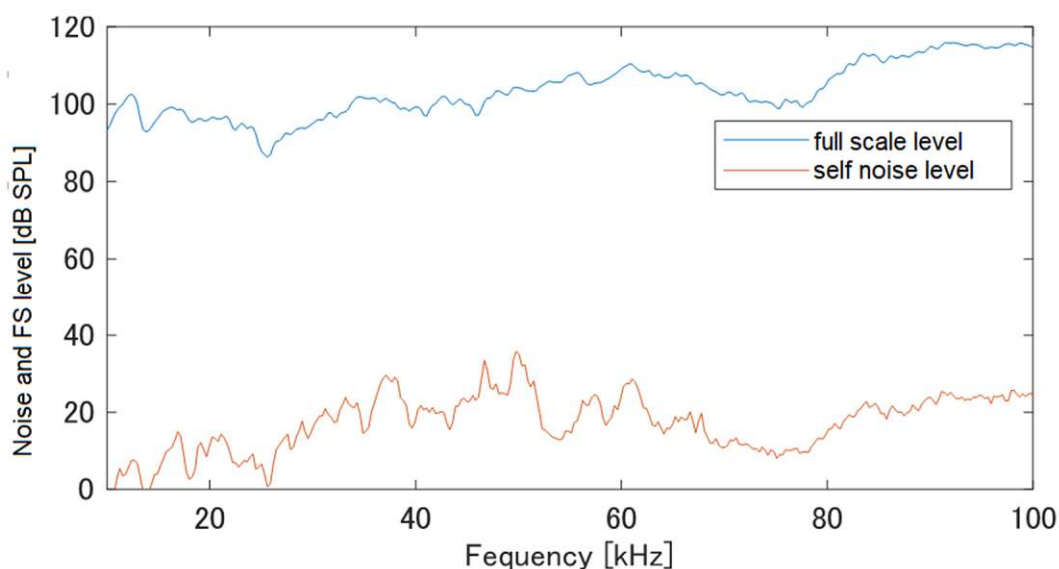

Figure 5: Recording characteristic of MEMS microphone on the on-board recording device.

The flight times per recording were 3-5 minutes for recordings made with the standing microphone and 1-3 minutes for recordings made with the on-board data logger. When recording with the logger, the bats flew alone before and after the pair flight.

2.3 Processing of data

The acquired sound files were cut automatically using software written in Python. The script detects a rise in the amplitude and then waits until it falls under a threshold. If the calls surpasses a minimum length and volume it is saved.

As described above, data loggers were used in pair flights to be sure of the true originator of each call. The issue was still not solved however, as with the data logger it could still happen that one of the attached loggers will pick up a call of the other bat. For this reason, the obtained pair flight recordings were synchronized and those calls recorded by the first logger were discarded, where the second logger had recorded calls within a set time window prior to the first one (fig. 6). Exclusion window lengths, t_{window}, used were 15 ms and 25 ms. As the echolocation calls emitted by one bat need time to propagate to (the attached microphone of) the other, these values are derived from the speed of sound, $c_0 = 343$ m/s, and the maximum distance, d_{max}, between the two bats, where one device can still "hear" the calls from the opposing bat.

$$d_{max} = c_0 \cdot t_{window}$$

While a t_{window} of 15 ms thus corresponds to a propagation distance within $d_{max} = 5.15$ m, a t_{window} of 25 ms corresponds to calls emitted within $d_{max} = 8.58$ m, larger than the flight space (fig. 3) diagonal of 7.87 m.

This method may exclude some calls that were truly those of the first bat, however it strictly excludes the calls that could have originated from the second one.

As and additional comparison, the synchronized recordings of the pair flights were cut manually in a way that only those segments were left were both individuals were emitting sounds simultaneously. Since *M. fuliginosus* only emits echolocation calls during flight, this corresponds to having only those calls left while both bats were airborne at the same time. This analysis was again done for exclusion window lengths of 15 and 25 ms.
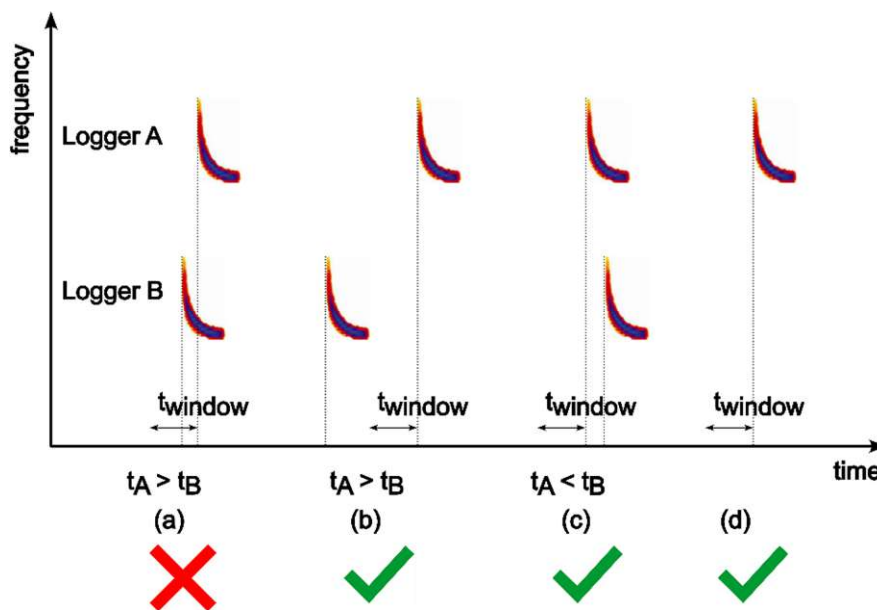

Figure 6: Call exclusions in pair flights. Two bats were equipped with on-board data loggers, recordings are synchronized. When interested in calls recorded with data logger A, only those calls are accepted where data logger B did not also record a call in a set time window before. (a) Logger B recorded a call temporally too close before logger A, so the call of A is rejected. (b) Logger B recorded a call before logger A, but outside of the exclusion window. (c) Logger B only recorded a call after logger A. (d) No calls were recorded by logger B before logger A.

After cutting, in another Python script, for each call a spectrogram was created. Additionally, a contrast function was used to eliminate any noise in the background and leave a clear image of the call and possible harmonics. After this, the spectrograms were looked at by myself to ensure no noise or other sounds were mistakenly seen as an echolocation call by the program. If images containing noise were found, they were deleted from the database.

8

Spectrograms were generated using the SciPy library's signal package. The length of the FFT was calculated via

$$n_{fft} = rate \,/\, cut\_freq * y\_res,$$

Where $rate$ is the sample rate of the recording, cut_freq is the upper frequency limit of the image and y_res is the resolution of the frequency range. As mentioned, the sampling rate of the standing microphone was 500 kHz and for the data logger it was 288 kHz. The upper frequency was set to 144 kHz, and the frequency resolution was set to 128 pixels for all images. The output image size was 512x128 pixels and the window type used was a Hanning window, with a window width of 128, in all cases.
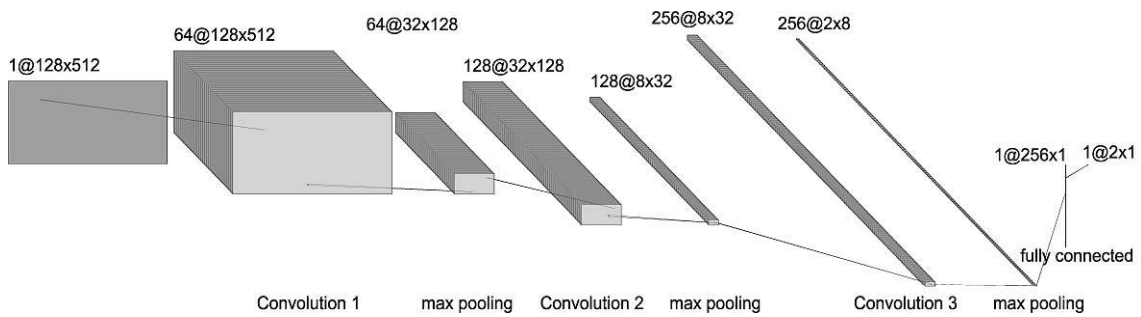
2.4 Neural network

Figure 7: Structure of the convolutional neural network used to classify the spectrograms. Not shown are the dropout layers after each convolutional and the final fully-connected layer. In the depicted case the outcome would be the probability of a call coming from 2 different individuals.

The structure of the neural network us ed in all analyses consisted of 3 convolutional layers with increasing channel sizes (64, 128 and 256 respectively) and using a ReLU (rectified linear unit) activation function, each followed by a max pooling layer (4x4 kernel) and a dropout layer ($p = 0.20$). Finally, two fully connected layers led to the outcome, the posterior probability for each class (bat individuals) that the input image (the spectrogram of an echolocation call) came from them (fig. 7). The PyTorch library [25] was used to develop, train and evaluate the CNN.

The outcome of the evaluation of a dataset can be summarized in a confusion matrix of observation and prediction (table 2).

Table 2: Confusion matrix of observation and prediction.

Confusion matrix		Prediction	
		Positive	Negative
Observation	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

The performance of the model was evaluated by calculating the F1 score, which itself can be calculated from precision, and recall of the model.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

While precision is a metric to answer the question "Of all positive predictions, how many are really positive?", recall is a metric to answer the question "Of all real positive cases, how many are predicted positive?". In other words, if one wants to avoid mainly false positives, they should improve their model according to precision, and if they want to avoid mainly false negatives, recall should be considered. If, on the other hand, both FP and FN should be avoided equally, the harmonic mean of precision and recall, the F1 score should be used as a metric, as was done in this case.

$$\text{F1 score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{TP}{TP + 0.5\,(FP + FN)}$$

With a confusion matrix showing how each call was matched to an individual, the F1 score could be calculated for each bat. All bats' individual (micro) F1 scores were then averaged to gain the macro F1 score as a final indicator for each experiment.

The NN was trained with 90% of the dataset and tested with the remaining 10% in a 10-fold setup. (fig. 8). This was repeated for 10 different random seeds and results were then averaged. Each training cycle consisted of 100 epochs. The reason for using a k-fold setup is to prevent overfitting, a situation where the model is trained so strongly on the training data, that it has difficulties predicting the correct labels for the testing data. When using k-folds, the model is eventually evaluated with all of the dataset, preventing this scenario.
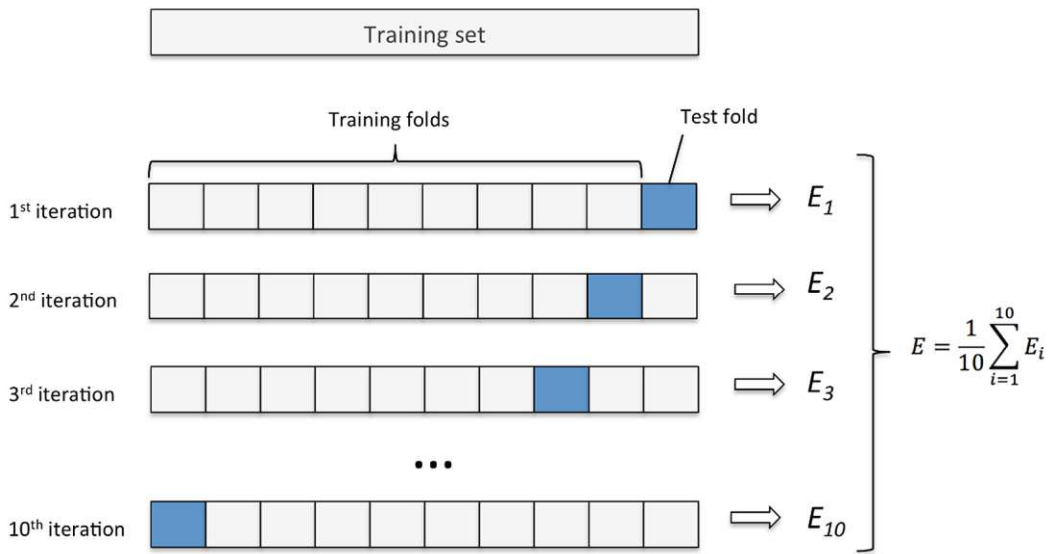
Figure 8: 10-fold cross validation. Each fold was evaluated with 10 random starting seeds. The outputs E_i for each iteration was the micro F1 score, which were than averaged to the macro F1 score.

For the analysis of bats flying alone, the processed recordings of all sessions in which the standing microphone was used, were combined. The calls of those individuals of which most calls were available were used to train and test the NN.

The analysis of the *M. fuliginosus* pair flight experiment was conducted with data logger recordings only. Each model was trained with the calls of the two bats flying alone and tested with the calls left over after call exclusion (chapter 2.3) of the bats flying together. For each pair four models were trained, they included all combinations of exclusion window times and whether only those calls were used where both bats were airborne simultaneously or not.

For the two pairs where there were enough calls left over even after close calls were excluded, a respective model was trained to compare the performances of solo-flight and pair-flight trained models.

2.5 Local Interpretable Model-Agnostic Explanations (LIME)

Usually, when using machine learning, the user does not know exactly how a model arrives at its decision and just has to trust on it's prediction. To overcome this black-box nature, an "explainer" can show the parameters of the input that are most important to the model. LIME has the advantage of being model-agnostic, meaning that the underlying model doesn't need to be known by it. The explanation is generated by approximating the model by an interpretable one, for example one that is linear and has only a few non-zero coefficients. The black-box model is then

estimated locally, in the neighbourhood of the prediction. For image classification, the image is divided into contiguous superpixels (fig. 9) which are turned on and off (blacked out) in different permutations. The simple, linear model is then trained with the probabilities for each permutation. Each superpixel is weighted in this process and those with the highest positive weights are chosen as the explanation (fig. 10). [11]

The image segmentation function used for all LIME analyses in this thesis was the Felzenswalb algorithm [26]. The size of the neighbourhood to learn the linear model was set to 100 and the 10 most important superpixels were plotted.
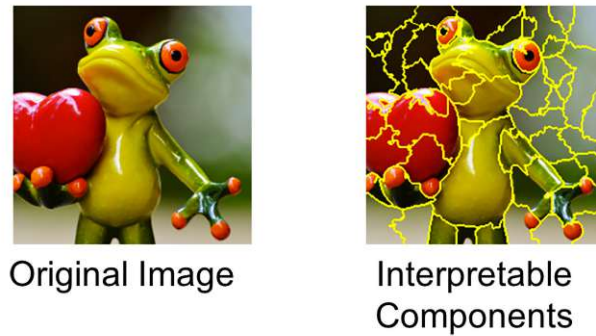

Figure 9: An image is split into superpixels by LIME. Image sources: Marco Tulio Ribeiro, Pixabay.
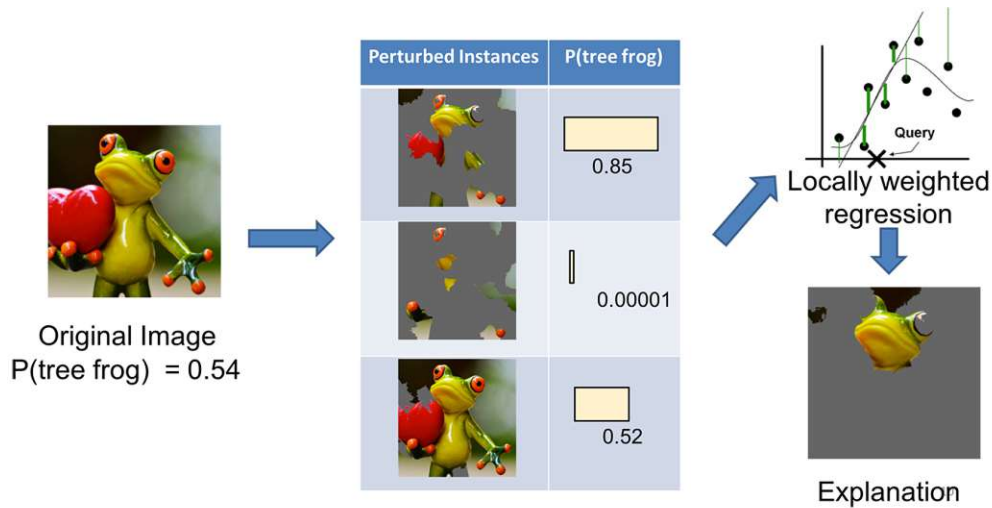

Figure 10: Different superpixels are turned on and off, those permutations yield data on which a simple, linear model is trained. Superpixels with the highest positive weights are then presented as the explanation. Image sources: Marco Tulio Ribeiro, Pixabay.

3 Results

3.1 Call collection

The total number of calls recorded with the standing microphone for *M. fuliginosus* and *P. abramus* flying alone are shown in table 3 and visualized in figure 11.

Table 3: Calls acquired for *M. fuliginosus* and *P. abramus* flying alone.

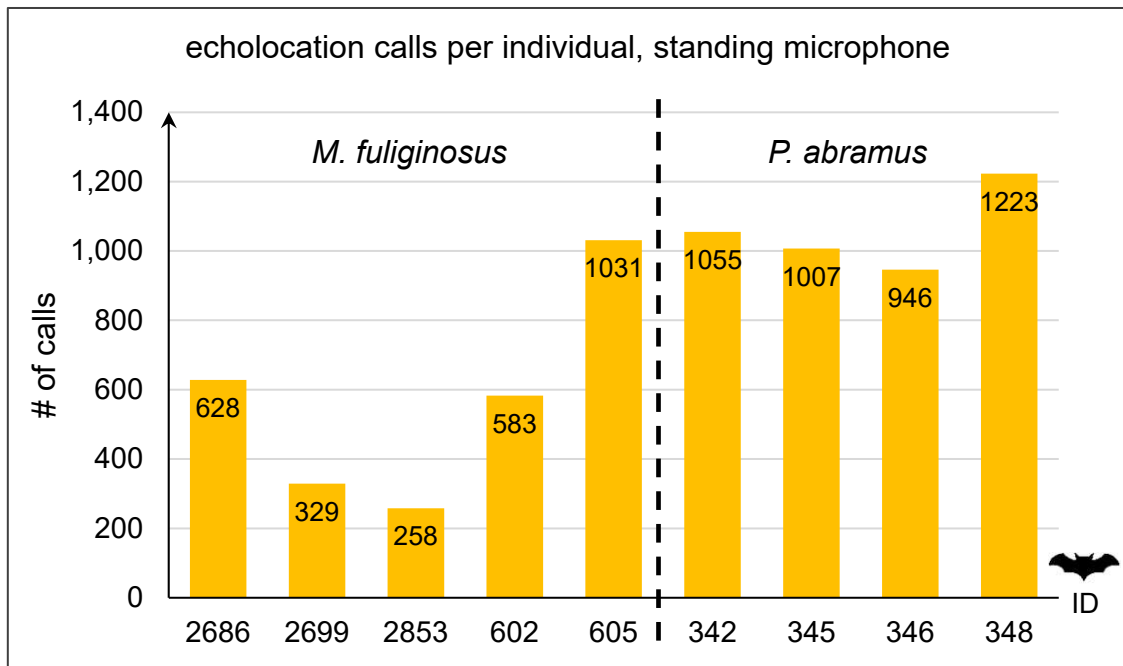| Bat Individual ID (*M. fuliginosus*)	# of calls	Bat Individual ID (*P. abramus*)	# of calls
2686	628	342	1,055
2699	329	345	1,007
2853	258	346	946
602	583	348	1,223
605	1,031		

Figure 11: Calls collected per individual bat in solo flight, recorded with a standing microphone.

The number of calls acquired for 4 pairs of *M. fuliginosus* flying alone (used for training) before and after the pair flight (used for testing), are shown in table 4 and visualized in figure 12. All of these recordings were performed with the onboard data logger. For each pair four different

13

datasets were created, differing in call exclusion windows lengths and if the used audio included only calls where both bats are airborne simultaneously or not.

Table 4: Datasets acquired for the pair flight experiments depending on call exclusion window. Calls for training were recorded during solo flight and calls for testing were recorded in pair flight.

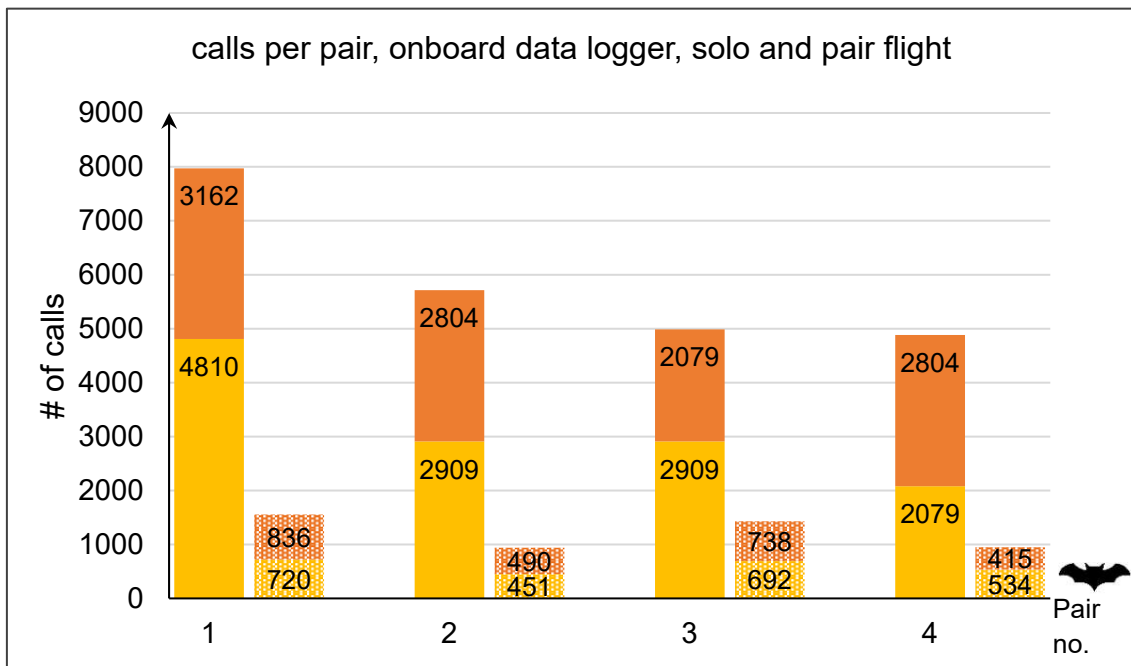| Pair no.	Bat Individual ID	# of calls, solo flight	# of calls for testing, 15 ms exclusion	# of calls for testing, 25 ms exclusion	# of calls for testing, 15 ms exclusion, movement only	# of calls for testing, 25 ms exclusion, movement only
1	605	4,810	1,344	1,240	843	720
	2686	3,162	1,142	1,064	901	836
2	601	2,909	1,165	1,080	539	451
	2690	2,804	698	649	535	490
3	601	2,909	1,129	1,012	816	692
	605	2,079	1,357	1,286	810	738
4	605	2,079	1,015	906	630	534
	2690	2,804	1,341	1,254	492	415

Figure 12: Calls collected for each pair of bats. For each pair the calls collected in solo flights are shown in the respective left (solid) column and the calls collected in pair flight are shown in the respective right (dotted) column. The pair flight echolocation calls represented are those, that are left over after using a 25 ms exclusion window and recordings of both bats flying simultaneously only.

3.2 Classification results

The F1 scores for the solo flight experiments are shown in table 5.

Table 5: Classification results for solo flight experiments.

Bat species	F1 score
P. abramus (standing microphone)	0.710
M. fuliginosus (standing microphone)	0.842
M. fuliginosus (on-board data logger)	0.924

The F1 scores for the pair flight experiments are shown in table 6 and compared to the solo flight experiment in figure 13. Scores for solo flight data for the respective pairs are shown for comparability. All individuals were of species *M. fuliginosus*. For each model trained on a pairs solo flight data, the 4 respective pair flight datasets created were evaluated.

Table 6: Classification results for pair flight experiments.

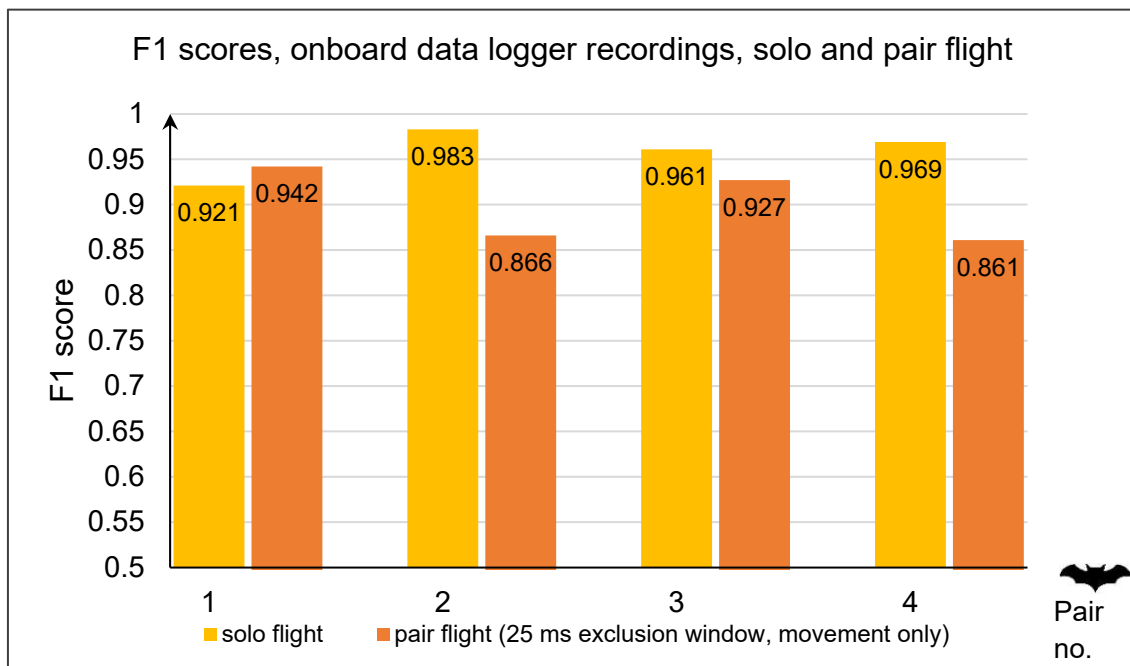| Pair no.	Bat Individual ID	F1 score				
		solo flight	pair flight, 15 ms exclusion	pair flight, 25 ms exclusion	pair flight, 15 ms exclusion, both airborne only	pair flight, 25 ms exclusion, both airborne only
1	605	0.921	0.946	0.943	0.947	0.942
	2686					
2	601	0.983	0.895	0. 899	0.868	0.866
	2690					
3	601	0.961	0.930	0.925	0.933	0.927
	605					
4	605	0.969	0.885	0.880	0.872	0.861
	2690					

Figure 13: Respective F1 scores in solo and pair flights for 4 pairs of bats.

For pairs no. 1 and 3 enough calls were obtained in pair-flight to *train* their own model. Their results are shown in the right-most column of table 7.

Table 7: Classification results for models trained with calls collected in pair flight (right-most column).

Pair no.	Bat Individual ID	Trained with solo-flight data		Trained with pair-flight data
		F1 score, solo flight	F1 score, pair flight, 25 ms exclusion, both airborne only	
1	605	0.921	0. 942	**0.948**
	2686			
3	601	0. 961	0. 927	**0.956**
	605			

3.3 Explanations by LIME algorithm

Examples of images generated with LIME are shown in figures 14 to 27. Representative images (spectrograms overlayed with LIME superpixels) have been plotted for the analyses of the solo flight data recorded of *M. fuliginosus* and *P. abramus* with the standing microphone when flying alone (table 5), as well as of the analyses of all *M. fuliginosus* pairs in table 6 and 7. For each analysis an image with the average value of all LIME generated images for each pixel has been

plotted as well. The values of all axes are pixels. Colors indicate the importance of each superpixel for the decision made by the model, encoded with the following colormap:

least important   most important
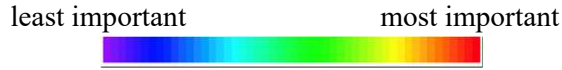
3.3.1 Standing microphone recordings

The following images were generated with calls represented in table 3 and models whose outcomes are represented in table 5.
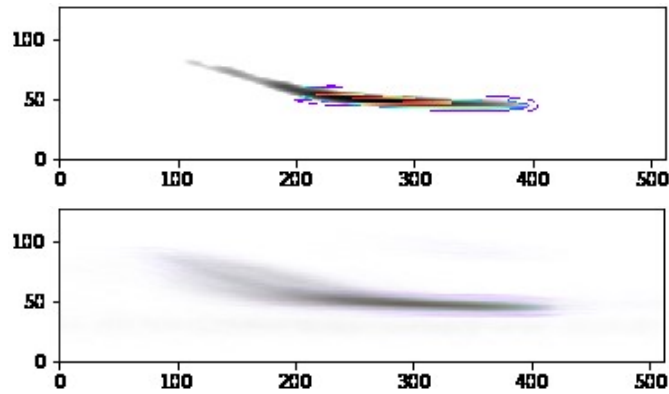

Figure 14: LIME analysis, representative correctly identified image of *M. fuliginosus* (bat ID 2686, top) and average image (bottom).
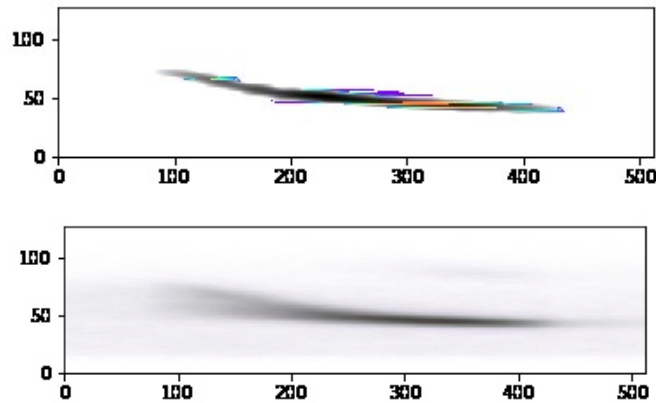

Figure 15: LIME analysis, representative correctly identified image of *P. abramus* (bat ID 345, top) and average image (bottom).

3.3.2 On-board data logger recordings

The following images were generated with calls represented in table 4 and models whose outcomes are represented in table 6. These models were trained with solo flight data and tested with pair flight data. Only the case of 25 ms exclusion time of movement-only data is used. Since these recordings were made with the on-board data logger, all following calls represented are from *M. fuliginosus*.
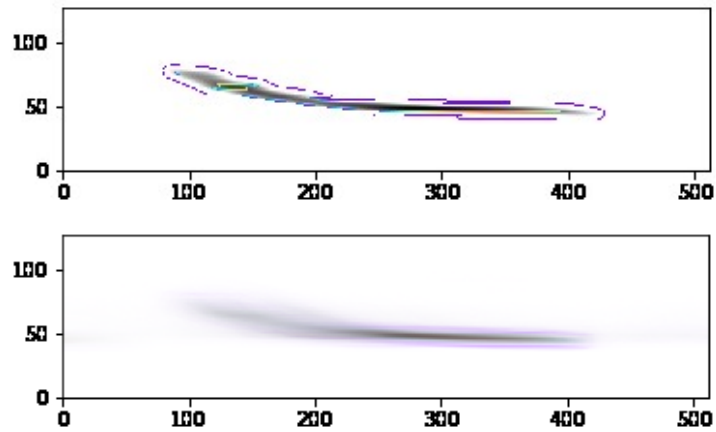
Figure 16: LIME analysis, pair 1. Representative call of bat 2686 correctly identified (top) and average image (bottom).
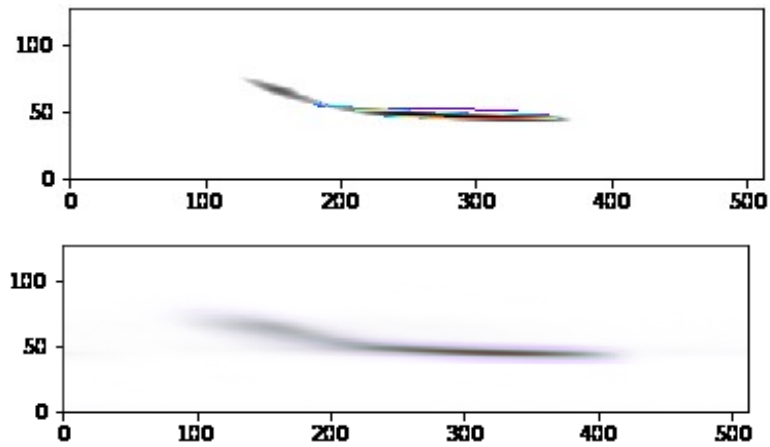

Figure 17: LIME analysis, pair 2. Representative call of bat 601 correctly identified (top) and average image (bottom).
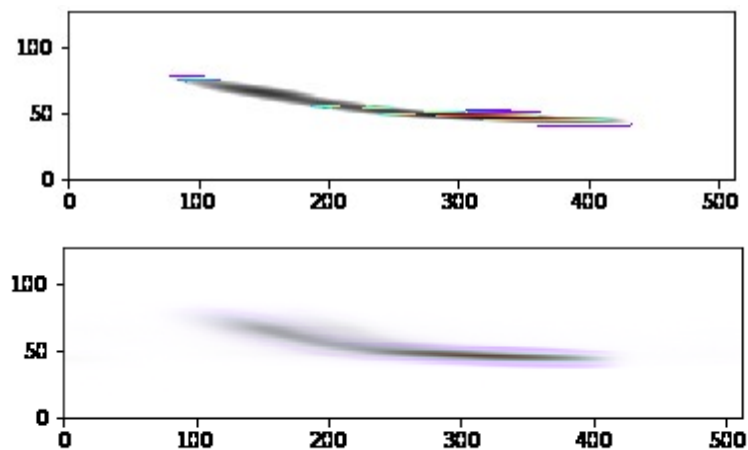

Figure 18: LIME analysis, pair 3. Representative call of bat 605 correctly identified (top) and average image (bottom).
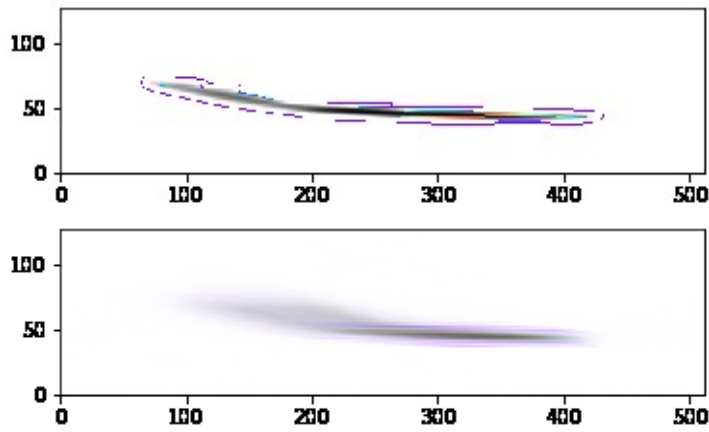
Figure 19: LIME analysis, pair 4. Representative call of bat 2690 correctly identified (top) and average image (bottom).

The next set of images were generated with the models whose outcomes are represented in table 7. These models were trained and tested with pair flight data.
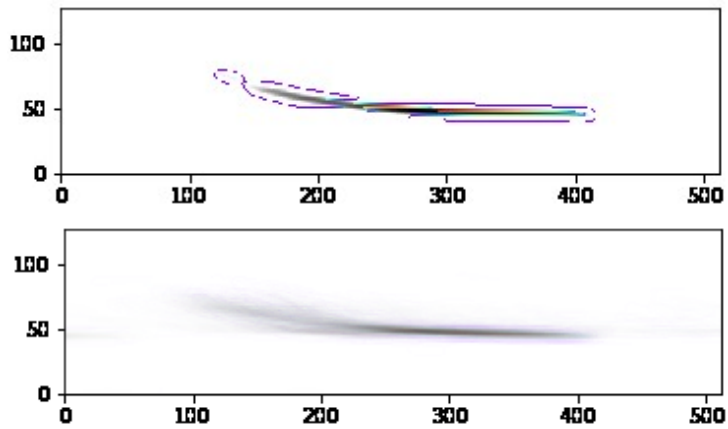

Figure 20: LIME analysis, pair 1. Representative call of bat 2686 correctly identified (top) and average image (bottom).
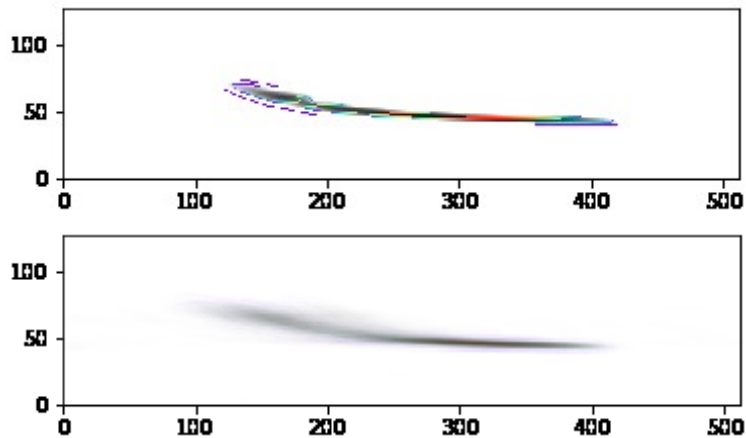

Figure 21: LIME analysis, pair 3. Representative call of bat 601 correctly identified (top) and average image (bottom).

3.3.3 Misidentified calls

Followingly examples of misidentified calls and their LIME analyses are presented.
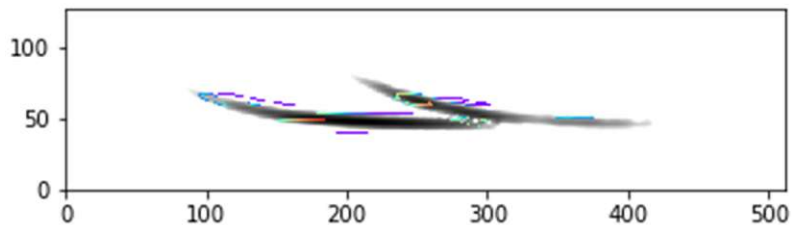
Figure 22: LIME analysis, standing microphone, solo flight, *M. fuliginosus*. Interpreted as bat ID 2699, while true caller was bat ID 2686.
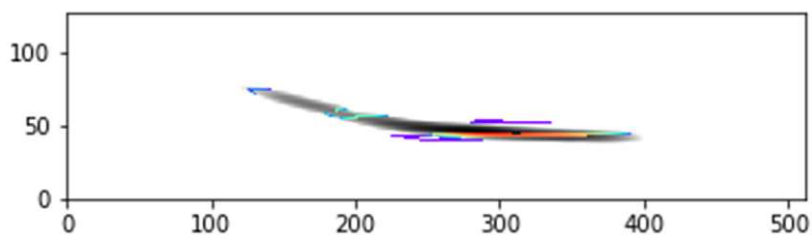
Figure 23: LIME analysis, standing microphone, solo flight, *P. abramus*. Interpreted as bat ID 346, while true caller was bat ID 342.

Figure 24: LIME analysis, pair 1. Interpreted as bat ID 2686, while true caller was bat ID 605.

Figure 25: LIME analysis, pair 2. Interpreted as bat ID 2690, while true caller was bat ID 601.
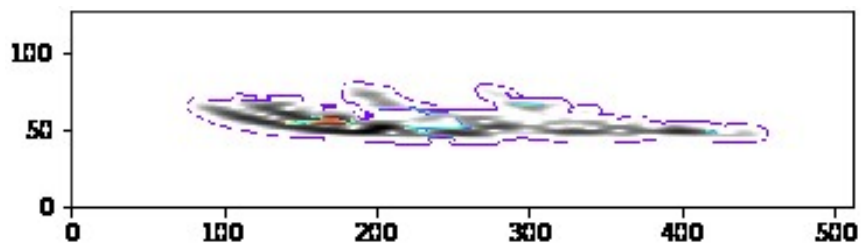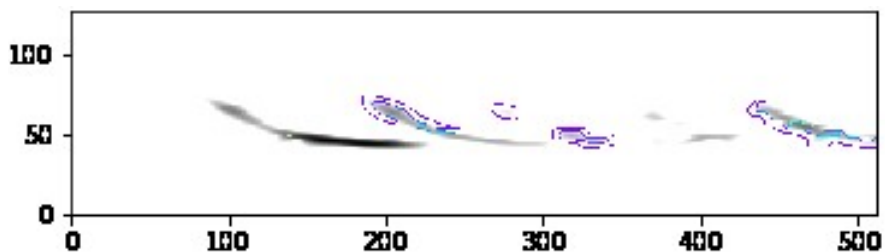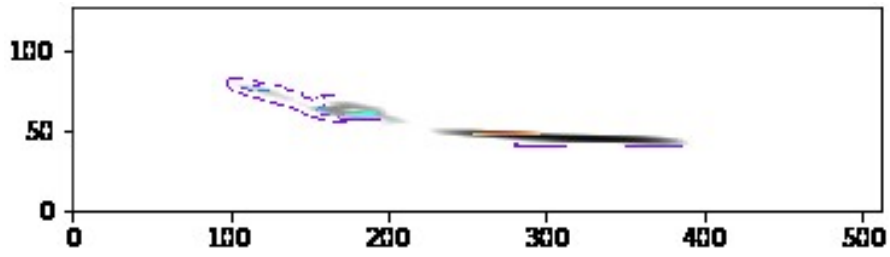
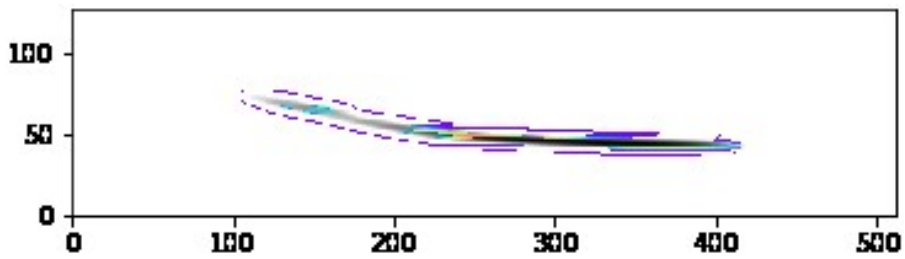Figure 26: LIME analysis, pair 3. Interpreted as bat ID 605, while true caller was bat ID 601.

Figure 27: LIME analysis, pair 4. Interpreted as bat ID 605, while true caller was bat ID 2690.

4 Discussion

4.1 Echolocation call classification

For the first time, conspecific bats could be distinguished by their echolocation calls alone in this study. These results indicate an individual-specific cue is contained within the calls.

In solo-flight recordings conducted with the standing microphone, the F1 score of analysis using the *M. fuliginosus* calls was 0.842 compared to 0.710 of *P. abramus* calls (table 5). Since the number of individuals was similar (5 and 4), this parameter is unlikely to have made a difference when comparing the two classification scores of the NN. Furthermore, the range of the number of calls per individual used to train the NN was much smaller for *P. abramus*, yet still leading to a lower F1 score. With this, I presume that the success rate is different mainly because of the fundamental difference of the echolocation calls between the two species: While *M. fuliginosus* bats have a more developed and longer TF part in their echolocation calls, *P. abramus* bats' calls may be shortened due to the flight setting in a confined space. [27] This different echolocation behaviour may also stem from the different habitats they live in in the wild. *P. abramus* bats don't live in large colonies in caves, as *M. fuliginosus*, and therefore have no need to develop significantly different call patterns between individuals; making it more difficult for the NN to tell them apart.

4.1.1 On-board data logger recordings, solo flight evaluation

For solo flight recordings of *M. fuliginosus* conducted with the on-board data logger, the F1 score was 0.924 for a model trained and evaluated with all 4 bats that used the device (table 5). Similar results were obtained when only the calls of the pair itself was used to train an, in this case binary, model. The respective models' F1 scores were, with a minimum of 0.921 for pair no. 1, also higher for all pairs (table 6, "solo flight" column), compared to the score of 0.842 for calls recorded with the standing microphone. An explanation for this increase performance might be the much larger call database collected with the data logger. Furthermore the calls of these databases do not include any Doppler shifts, i.e. are less distorted than when recorded with a standing microphone.

4.1.2 Pair flight evaluation, solo flight model

When evaluating the solo flight trained models with calls recorded in pair flight (table 6, figure 13), the F1 scores dropped when compared to the same trained models evaluated with the solo flight calls in 3 out of 4 pairs, with pair no. 1 as an exception (0.921 for solo flight calls vs. 0.942 for pair flight calls [25 ms exclusion window, movement only data]). This may well be an indication of jamming avoidance behaviour by the bats: as the model has more difficulty distinguishing between the same bats, with the only difference being the calls stemming from a pair flight and not solo flight scenario, one can assume the reason for this is due to the bats

changing their individual echolocation calls to avoid interference of their respective echoes. The call characteristic that has been shown to change most significantly in previous studies is the terminal frequency of the calls. [10] This observation is also fitting with the analysis made with LIME in this thesis (see below).

Furthermore, the F1 scores were lower when increasing the exclusion window size in all instances except for pair no. 2 when using the entire recordings (calls where one bat was not airborne included). This indicates that unnecessarily many calls had been excluded for the 25 ms window and due to the reduced dataset caused a worse performance.

Comparing the results for using only calls recorded when both bats are airborne simultaneously, versus when also including calls recorded when one of the bats was temporarily resting, the results are ambiguous and improve or diminish almost equally as often. An explanation for this might be that the bats sometimes keep the adaption of the TF of their echolocation calls even when the opposite bat is currently not emitting any pulses.

4.1.3 Pair flight evaluation, pair flight model

Two pairs (pair no. 1 and 3) had models trained also with pair flight calls (table 7), and in both cases the F1 score improved when compare to the model having been trained with calls from solo flights. This, of course, makes sense as in the improved version the training dataset stemmed from the same recording session as the testing dataset. This would actually be the best case for a NN that needs to distinguish calls. However, the minimal difference (0.027 for pair no.1 and 0.005 for pair no. 3) to the F1 scores obtained with models trained and tested with solo flight calls, suggests, that the possible changes of the bats' echolocation calls are not big enough to significantly reduce the performance of the neural network. This is a positive sign for a possible future system that, set up in nature, will need to distinguish many bats flying at once, with each of them changing their TF due to the group flight situation.

4.2 LIME analysis

Spectrograms overlaid with the LIME superpixels (fig. 14 to 21), the relevant image areas for the NN's decision, show a clear pattern. In most cases of correct identification the TF part of the calls' spectrogram is the most important feature to determine the individual. This result was observed across both species, in solo and pair flight, as well as all combinations of call exclusion criteria, and can therefore be linked to previous studies like [10], indicating the individual characteristic of the TF of echolocation calls of bats.

The average images of all respective LIME generated images for each analyzed dataset show a similar result, with the brightest spots being the QC-TF parts of the call. Furthermore, by focusing just on the black color parts, the echolocation calls themselves, the uniformity of the calls is

visible. For example, in figure 15 (bottom), showing *P. abramus* calls, there seems to be more noise and/or less uniformity in the spectrograms than in figure 14 (bottom).

Another practical use of LIME is to find out why wrong decisions were made by the NN. Figures 22, 24 and 25 show spectrograms where more than one call or the initial call and an echo are depicted. As the model was trained on images almost exclusively depicting only one call without any echoes, the NN will of course have more trouble identifying these deviating cases. Figure 26 shows the case of a call to weak in amplitude in parts, so that the spectrograms was not bright enough to be recognized. Figures 23 and 27 show proper spectrograms, and in these cases it can be assumed the calls where just too similar to those of the bat's it was mistaken for.

4.3 Call collection

Recordings conducted of bats flying solo and recorded with a standing microphone in 7 recording sessions yielded a range from 258 to 1,031 usable calls (i.e. calls after processing of audio recordings) for 5 *M. fuliginosus* individuals and a range of 946 to 1,223 calls for 4 *P. abramus* individuals (table 3, figure 11). Much more usable calls were able to be retrieved when recording with the on-board data logger, where with just one recording session 2,079 to 4,810 usable calls per bat in solo flight were retrieved (table 4). This may well be explained by the difference in recording location of the different microphones. A standing microphone will not pick up a bat's call when the distance between sender and receiver is too big and the call is attenuated too much as it propagates in the air. This is not the case when the call is recorded directly after emission as with the data logger. The improvement in recording calls seems to also outweigh the reduced recording time with the logger due to increased strain of the bat while flying with a device that is over 20 % of their body weight (table 1).

For calls recorded of *M. fuliginosus* flying in pairs, some more observations can be made. Firstly, the number of calls is much less than in solo flight, with a maximum of 1,357 calls for bat ID 605 in pair 3, even when only applying the weakest selection criteria (15 ms exclusion window and data of one bat resting included, table 4). Secondly, when increasing the call exclusion window from 15 ms to 25 ms, the calls left over are reduced, as with a bigger window length more calls will fall into the exclusion criteria. Thirdly, when reducing the recordings to only those times where both animals were airborne at the same time, again less calls were extracted as the leftover sound data was shorter.

5 Conclusion and outlook

With the experiments and analyses described in this thesis, I showed for the first time that, individual, conspecific bats of the two species *Miniopterus fuliginosus* and *Pipistrellus abramus*, can be distinguished by their echolocation calls alone. The individual specific signature was found by LIME analysis to most likely be the constant terminal frequency part of the echolocation calls which, indicated by reduced performance of the NN and as suggested in previous studies, might change by an individual bat when not flying alone (jamming avoidance behaviour). Whether the bats themselves can tell each other apart by eavesdropping on the echolocation calls of their conspecifics is not totally clear, but was suggested in previous studies. [12]–[14]

In terms of audio recording, overall, the on-board data logger proved to be a more successful device to record calls of bats than the standing microphone, as attenuations and possible Doppler-shifts are not included in its recordings. However, it has to be noted that for a setting of an automated system in nature this is not feasible. It will, nevertheless, be a useful tool in further experiments in the future and give an insight to the bats' hearing perspective during flight.

A possible improvement in individual classification could lie in the processing step of the audio recordings. There, noise that was mistaken for echolocation calls by the Python script that cut the audio file, was deleted by hand. Instead, it could be useful to take these cuts and put them into a separate class labelled "noise" or "unknown". This way the NN could be trained to also identify noise, or other sounds that are not echolocation calls, automatically.

As a first application in the wild, microphones could be set up in nature close to a known feeding area. By recording the bats' calls some conclusion might be drawn into whether it is always the same individual(s) returning or if there is more variety. Furthermore, such an experiment could be conducted without needing to catch and tag any animals.

6 Acknowledgements

I would like to thank my fellow students at the Neuroethology and Bioengineering Laboratory of Doshisha University for taking me in with such hospitality, even though communication through the language barrier was not always easy. Soshi Yoshida was especially helpful, not just for helping me with the research and experiments I conducted, but also by helping me with countless tasks related to my daily life in Japan. Furthermore I would like to thank Professor Shizuko Hiryu of this laboratory for her academic guidance and support in all of my research, publishing and this thesis. I am also grateful to Olga Heim for getting me started in my research topic as well as helping me with publishing my findings, and to Professor Keisuke Imoto of the Faculty of Science and Engineering, Doshisha University, for his help in refining the neural network used for my analyses.

From my home university TU Wien, I want to give special thanks to Professor Eugenijus Kaniusas for his inspiring lectures, and together with Mr. Thomas Rief, for their always prompt help with questions regarding my double degree.

Finally, my gratitude goes to my family for supporting me from near and far.

Thank you very much.

Vielen Dank!

ありがとうございました.

7 References

[1]   G. Neuweiler, *The Biology of Bats*. Oxford University Press, 2000.

[2]   N. B. Simmons, "Order Chiroptera," in *Mammal Species of the World: a Taxonomic and Geographic Reference, vol. 1*, 2005, pp. 312–529.

[3]   H. Raghuram and M. Jain, "Species and acoustic diversity of bats in a palaeotropical wet evergreen forest in southern India," 2014. [Online]. Available: https://www.researchgate.net/publication/265551050

[4]   K. Kobayashi, K. Masuda, C. Haga, T. Matsui, D. Fukui, and T. Machimura, "Development of a species identification system of Japanese bats from echolocation calls using convolutional neural networks," *Ecol Inform*, vol. 62, May 2021, doi: 10.1016/j.ecoinf.2021.101253.

[5]   J. Rydell, S. Nyman, J. Eklöf, G. Jones, and D. Russo, "Testing the performances of automated identification of bat echolocation calls: A request for prudence," *Ecol Indic*, vol. 78, pp. 416–420, Jul. 2017, doi: 10.1016/j.ecolind.2017.03.023.

[6]   B. M. Siemers and G. Kerth, "Do echolocation calls of wild colony-living Bechstein's bats (Myotis bechsteinii) provide individual-specific signatures?," *Behav Ecol Sociobiol*, vol. 59, no. 3, pp. 443–454, Jan. 2006, doi: 10.1007/s00265-005-0068-x.

[7]   R. D. Redgwell, J. M. Szewczak, G. Jones, and S. Parsons, "Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks," *Algorithms*, vol. 2, no. 3, pp. 907–924, Sep. 2009, doi: 10.3390/a2030907.

[8]   X. Chen, J. Zhao, Y. hua Chen, W. Zhou, and A. C. Hughes, "Automatic standardized processing and identification of tropical bat calls using deep learning approaches," *Biol Conserv*, vol. 241, p. 108269, Jan. 2020, doi: 10.1016/J.BIOCON.2019.108269.

[9]   J. T. Treitler, O. Heim, M. Tschapka, and K. Jung, "The effect of local land use and loss of forests on bats and nocturnal insects," *Ecol Evol*, vol. 6, no. 13, pp. 4289–4297, Jul. 2016, doi: 10.1002/ece3.2160.

[10]   K. Hase, Y. Kadoya, Y. Maitani, T. Miyamoto, K. I. Kobayasi, and S. Hiryu, "Bats enhance their call identities to solve the cocktail party problem," *Commun Biol*, vol. 1, no. 1, Dec. 2018, doi: 10.1038/s42003-018-0045-3.

[11]   M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*, pp. 97–101, Feb. 2016, doi: 10.18653/v1/n16-3020.

[12] Y. Yovel, M. L. Melcon, M. O. Franz, A. Denzinger, and H. U. Schnitzler, "The voice of bats: How greater mouse-eared bats recognize individuals based on their echolocation calls," *PLoS Comput Biol*, vol. 5, no. 6, 2009, doi: 10.1371/journal.pcbi.1000400.

[13] M. Schuchmann, S. J. Puechmaille, and B. M. Siemers, "Horseshoe bats recognise the sex of conspecifics from their echolocation calls," *Acta Chiropt*, vol. 14, no. 1, pp. 161–166, Jun. 2012, doi: 10.3161/150811012X654376.

[14] G. Jones and B. M. Siemers, "The communicative potential of bat echolocation pulses," *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, vol. 197, no. 5. pp. 447–457, May 2011. doi: 10.1007/s00359-010-0565-x.

[15] W. Mitch Masters, K. A. S Raver, K. A. Kazial, and W. M. Masters, "Sonar signals of big brown bats, Eptesicus fuscus, contain information about individual identity, age and family affiliation," 1995.

[16] S. C. Burnett, K. A. Kazial, and W. M. Masters, "Discriminating individual big brown bat (eptesicus fuscus) sonar vocalizations in different recording situations," *Bioacoustics*, vol. 11, no. 3, pp. 189–210, Jan. 2001, doi: 10.1080/09524622.2001.9753462.

[17] K. Hase, T. Miyamoto, K. I. Kobayasi, and S. Hiryu, "Rapid frequency control of sonar sounds by the FM bat, Miniopterus fuliginosus, in response to spectral overlap," *Behavioural Processes*, vol. 128, pp. 126–133, Jul. 2016, doi: 10.1016/j.beproc.2016.04.017.

[18] Y. Maitani, K. Hase, K. I. Kobayasi, and S. Hiryu, "Adaptive frequency shifts of echolocation sounds in Miniopterus fuliginosus according to the frequency-modulated pattern of jamming sounds," *Journal of Experimental Biology*, vol. 221, no. 23, Dec. 2018, doi: 10.1242/jeb.188565.

[19] T. Furuyama, K. Hase, S. Hiryu, and K. I. Kobayasi, " Hearing sensitivity evaluated by the auditory brainstem response in Miniopterus fuliginosus ," *J Acoust Soc Am*, vol. 144, no. 5, pp. EL436–EL440, Nov. 2018, doi: 10.1121/1.5079904.

[20] E. Takahashi, K. Hyomoto, H. Riquimaroux, Y. Watanabe, T. Ohta, and S. Hiryu, "Adaptive changes in echolocation sounds by Pipistrellus abramus in response to artificial jamming sounds," *Journal of Experimental Biology*, vol. 217, no. 16, pp. 2885–2891, Aug. 2014, doi: 10.1242/jeb.101139.

[21] K. Motoi, M. Sumiya, E. Fujioka, and S. Hiryu, " Three-dimensional sonar beam-width expansion by Japanese house bats (Pipistrellus abramus) during natural foraging ," *J Acoust Soc Am*, vol. 141, no. 5, pp. EL439–EL444, May 2017, doi: 10.1121/1.4981934.

[22] S. Hiryu, T. Hagino, H. Riquimaroux, and Y. Watanabe, " Echo-intensity compensation in echolocating bats (Pipistrellus abramus) during flight measured by a telemetry

microphone ,” *J Acoust Soc Am*, vol. 121, no. 3, pp. 1749–1757, Mar. 2007, doi: 10.1121/1.2431337.

[23]   S. Hiryu, T. Hagino, E. Fujioka, H. Riquimaroux, and Y. Watanabe, “ Adaptive echolocation sounds of insectivorous bats, Pipistrellus abramus , during foraging flights in the field ,” *J Acoust Soc Am*, vol. 124, no. 2, pp. EL51–EL56, Aug. 2008, doi: 10.1121/1.2947629.

[24]   S. Boku, H. Riquimaroux, A. M. Simmons, and J. A. Simmons, “ Auditory brainstem response of the Japanese house bat (Pipistrellus abramus) ,” *J Acoust Soc Am*, vol. 137, no. 3, pp. 1063–1068, Mar. 2015, doi: 10.1121/1.4908212.

[25]   A. Paszke *et al.*, “Automatic differentiation in PyTorch.”

[26]   P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient graph-based image segmentation,” *Int J Comput Vis*, vol. 59, no. 2, pp. 167–181, Sep. 2004, doi: 10.1023/B:VISI.0000022288.19776.77/METRICS.

[27]   S. Hiryu, T. Hagino, E. Fujioka, H. Riquimaroux, and Y. Watanabe, “ Adaptive echolocation sounds of insectivorous bats, Pipistrellus abramus , during foraging flights in the field ,” *J Acoust Soc Am*, vol. 124, no. 2, pp. EL51–EL56, Aug. 2008, doi: 10.1121/1.2947629.